Psychological Assessment
2003, Vol. 15, No. 3, 413–425

1040-3590/03/$12.00 DOI: 10.1037/1040-3590.15.3.413

A Multisite Comparison of Actuarial Risk Instruments for Sex Offenders

Grant T. Harris and Marnie E. Rice
Penetanguishene Mental Health Centre

Vernon L. Quinsey
Queen's University

Martin L. Lalumière
Centre for Addiction and Mental Health and University of Toronto

Douglas Boer
Correctional Service of Canada

Carol Lang
Penetanguishene Mental Health Centre

Four actuarial instruments for the prediction of violent and sexual reoffending (the Violence Risk Appraisal Guide [VRAG], Sex Offender Risk Appraisal Guide [SORAG], Rapid Risk Assessment for Sex Offender Recidivism [RRASOR], and Static-99) were evaluated in 4 samples of sex offenders (N = 396). Although all 4 instruments predicted violent (including sexual) recidivism and recidivism known to be sexually motivated, areas under the receiver operating characteristic (ROC) were consistently higher for the VRAG and the SORAG. The instruments performed better when there were fewer missing items and follow-up time was fixed, with an ROC area up to .84 for the VRAG, for example, under such favorable conditions. Predictive accuracy was higher for child molesters than for rapists, especially for the Static-99 and the RRASOR. Consistent with past research, survival analyses revealed that those offenders high in both psychopathy and sexual deviance were an especially high-risk group.

The last 2 decades have seen great progress in the prediction of recidivism and violence among a variety of criminal and psychiatric populations (Monahan et al., 2001; Quinsey, Harris, Rice, & Cormier, 1998). Among sex offenders in particular, there has been an exponential increase in research on the prediction of general, violent, and sexual recidivism. For example, a PsycINFO journal article abstract search for the words *predict** (or *recid**) and *sex offen** revealed 19 articles between 1890 and 1969, 14 during the 1970s, 55 in the 1980s, and 114 in the 1990s (e.g., Rice, Harris, & Quinsey, 1990; Rice, Quinsey, & Harris, 1991). Part of the impetus for this research has been public safety concerns posed by repeat sex offenders, which has in turn led to new laws pertaining to these offenders requiring mental health professionals to make predictions about sex offenders' future criminal behavior. Another impetus for the burgeoning literature was the success of actuarial methods of prediction. A number of studies have reported large effect sizes in the prediction of violent and sexual recidivism for sex offenders (e.g., Rice & Harris, 1997).

Another recent advance in the field of violence prediction was the application of measures derived from receiver operating characteristics (ROCs; first applied to signal detection over 50 years ago; Swets, Dawes, & Monahan, 2000) to help investigators think more clearly about the task (Rice & Harris, 1995). Specifically, the area under the ROC is a measure of predictive accuracy that is conceptually and numerically equivalent to the common language effect size (McGraw & Wong, 1992; see Rice & Harris, 1995). An advantage of using ROC area as a measure of predictive accuracy rather than correlation or other measures such as sensitivity, specificity, or positive or negative predictive value is that its value is independent of the base rate of recidivism in the sample at hand and thus provides a way to compare the accuracy of different instruments developed on samples with different base rates of recidivism. The use of the ROC area as a measure of predictive accuracy, combined with a method of assigning relative costs to false positives versus false negatives, would allow for practitioners to determine optimum cut scores for the particular situation (Rice & Harris, 1995). In addition, ROC area provides a measure that allows for comparison of effect sizes across different fields of prediction. The many advantages of using ROC area as a general measure of accuracy for diagnostic and predictive purposes have led to a call for its use as the standard measure (Swets et al., 2000).

The Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993) was developed for the prediction of violent (including all hands-on sex offenses) recidivism among offenders and mentally disordered offenders. The instrument was developed on a sample of 618 men (approximately 15% of whom were sex offenders) assessed in a Canadian maximum-security hospital. About

Grant T. Harris, Marnie E. Rice, and Carol Lang, Research Department, Penetanguishene Mental Health Centre, Penetanguishene, Ontario, Canada; Vernon L. Quinsey, Department of Psychology, Queen's University, Kingston, Ontario, Canada; Martin L. Lalumière, Law and Mental Health Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada, and Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; Douglas Boer, Pacific Region, Correctional Service of Canada, Agassiz, British Columbia, Canada.

This research was supported by a research grant from the Ontario Mental Health Foundation. We thank Garth Coleman, Catherine Cormier, Sonja Dey, Melanie Gates, and Tina Smith Krans for coding the data. We also thank Zoe Hilton and Michael Seto for helpful comments.

Correspondence concerning this article should be addressed to Grant T. Harris, Penetanguishene Mental Health Centre, 500 Church Street, Penetanguishene, Ontario L9M 1G3, Canada. E-mail: gharris@mhcp.on.ca

half of the men were subsequently convicted and sentenced to prison, whereas most of the other half were treated in a secure psychiatric hospital. All of the men were later released, with an average opportunity for violent recidivism of 7 years. Several steps were taken in the development of the VRAG to ensure that it would perform well on cross-validation (Harris et al., 1993).

The ROC area for the VRAG on its construction sample was .76, equivalent to a large effect size by conventional standards (equating to a Cohen's *d* of 1.00; see Rice & Harris, 1995). Of even more importance, the ROC area for predictive accuracy in a cross-validation sample of child molesters and rapists was .77 (Rice & Harris, 1997). The instrument has since been found to significantly predict violent recidivism in over a dozen independent samples (e.g., Glover, Nicholson, Hemmati, Bernfeld, & Quinsey, 2002; Harris, Rice, & Cormier, 2002; for a complete list, go to http://www.mhcp-research.com/ragreps.htm). There also have been several studies supporting the accuracy of the VRAG for predicting recidivism among sex offenders (e.g., Barbaree, Seto, Langton, & Peacock, 2001; Bélanger & Earls, 1996; Quinsey, Book, & Skilling, 2003; Sjöstedt & Langström, 2002).

The success of the VRAG in predicting recidivism (general, violent, and sexual) among sex offenders led to its use as a basis for a special instrument to predict violent recidivism for sex offenders specifically. The Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, et al., 1998) was constructed by incorporating variables (e.g., deviant sexual preferences and history of sex offenses) shown to be important for sex offenders. Scoring instructions and normative data are provided in Quinsey, Harris, et al. (1998). The SORAG has 14 items, 10 of which are identical to VRAG items, and, thus, the SORAG is highly correlated with the VRAG (Barbaree et al., 2001, reported a correlation of .90). The SORAG, like the VRAG, was developed to predict new arrests or convictions for violent (including sexual) offenses. Several studies with sex offenders have shown it to have high accuracy (median ROC area = .75) in the prediction of violent (including sexual) recidivism and moderate accuracy in predicting offenses known to be sexual (Barbaree et al., 2001; Dempster, Hart, & Boer, 2001; Nunes, Firestone, Bradford, Greenberg, & Broom, 2002; Rice & Harris, 2002).

The Rapid Risk Assessment for Sex Offender Recidivism (RRASOR; Hanson, 1997) was developed using an initial item pool of seven easily scored variables found to predict sex offense recidivism in a meta-analysis (Hanson & Bussière, 1998). Through multiple regression and examining intercorrelations among the variables in each of seven different data sets, the number of variables was reduced to four, and a weighting system was determined. Several studies have shown the RRASOR to be a strong predictor of recidivism known to be sexually motivated (median ROC area = .73; Barbaree et al., 2001; Dempster et al., 2001; Sjöstedt & Langström, 2002; see also Hanson & Harris, 2000).

The Static-99 (Hanson & Thornton, 2000) was constructed by combining the RRASOR and an unpublished nonactuarial instrument. The scale was tested on four samples of sex offenders (the same offenders used to derive the RRASOR) and showed moderate to high accuracy in the prediction of both sexual and violent recidivism. Subsequent research has reported significant correlations between recidivism and scores on the Static-99 (median ROC area = .70; Barbaree et al., 2001; Nunes et al., 2002).

In recent studies, researchers have compared the predictive accuracy of several instruments using a single sample of offenders. Barbaree et al. (2001) compared the four instruments above plus two others in a sample of 215 sex offenders followed for an average of 4.5 years. The four instruments used in the present study performed better than the other two, and there were no significant differences among those four. All were moderate to large predictors of violent recidivism, with the SORAG yielding the highest predictive accuracy (ROC area = .73), and all significantly predicted sexual recidivism. Sjöstedt and Langström (2002) compared the VRAG, RRASOR, Psychopathy Checklist—Revised (PCL–R), and a nonactuarial checklist in the prediction of recidivism among 51 rapists in Sweden. They reported that only the VRAG and PCL–R yielded statistically significant correlations with violent recidivism and only the RRASOR was significantly correlated with recidivism known to be sexually motivated.

In addition to the moderate to high predictive accuracies obtained using actuarial instruments, survival analyses have suggested that the combination of psychopathy and sexual deviance is a very good predictor of violent and/or sexual recidivism. Rice and Harris (1997) found an interaction of psychopathy and sexual deviance such that offenders high on both factors were charged or convicted for new offenses known to be sexual at a rate faster than that predicted by the additive effects of each variable alone, and the survival analysis showed that half of the sexually deviant psychopathic individuals committed new, clearly sexual offenses within 3 years of opportunity. A similar finding was obtained for general recidivism by Gretton, McBride, Hare, O'Shaughnessy, and Kumka (2001), whose study of juvenile sex offenders showed that sexually deviant psychopathic adolescent offenders reoffended at much higher rates than offenders who had only one or none of the two risk factors.

The purpose of the present study was to compare the accuracies of the four actuarial instruments in predicting the violent and sexual recidivism of male sex offenders. Typically, previous comparisons have involved sex offenders from a single institution (e.g., Barbaree et al., 2001); thus, we examined the predictive accuracy of these four instruments among four diverse samples of sex offenders from three different sites. We examined the performance of the instruments for rapists and child molesters separately. In addition, we examined the effect of the combination of psychopathy and sexual deviance.

Method

Overview

The data collection was entirely archival. The key features of the design were the union of historical data and specialized phallometric assessment data and independently coded recidivism data. We studied 396 male sex offenders, all of whom had offended against a child (child molesters; n = 170), an adult female (rapists; n = 191), or both (n = 35). Most had been incarcerated or hospitalized in one of three secure institutions, but 87 men were institutionalized for only very short periods, if at all, and resided in the community at the time of their original assessments. We evaluated the performance of four actuarial instruments: the VRAG, the SORAG, the RRASOR, and the Static-99. The coding of all predictor variables was conducted on the basis of archival material gathered before offenders received the opportunity to recidivate and was, therefore, truly predictive.

Participants

All of the men followed had been charged with a criminal offense involving sexual contact with a child under 15 years of age while the offender was at least 5 years older than the victim (child molesters), forceful or coercive sexual contact with an adult woman (rapists), or both. None had been included in previous studies on which the actuarial instruments had been developed. There were four groups. The first group (Oak Ridge) comprised all 118 sex offenders admitted to Oak Ridge and assessed in the Sexual Behaviour Laboratory, Penetanguishene, Ontario, Canada, between 1974 and 1994 who had not been included in earlier follow-up studies and who had an opportunity to reoffend before April 1, 1996. The majority had been admitted for psychiatric evaluation only and subsequently served sentences in federal or provincial correctional institutions.

The second group (community) comprised all 87 men assessed in the Oak Ridge Sexual Behaviour Laboratory from 1979 to 1994 who were referred from community sources (primarily provincial probation officials or federal parole officers) and who were at risk to reoffend at the time of the assessment. The third group (Kingston) comprised 96 federal inmates released from the Regional Treatment Centre, Kingston Penitentiary, Kingston, Ontario, Canada, between 1977 and 1989. These men were a randomly selected subsample of the participants reported elsewhere (Quinsey, Khanna, & Malcolm, 1998). In addition to variables already coded for that study, several additional variables were coded for the present study to score the VRAG and the SORAG. Finally, the fourth group (Pacific) included 95 released federal inmates, not reported on elsewhere, from the Regional Psychiatric Centre, Abbotsford, British Columbia, Canada, and released between 1978 and 1984.

Variables

The individual risk-related variables comprising the various actuarial tools are shown in Table 1. Most variables are self-explanatory, but those that require further explanation (e.g., the 20-item PCL–R; Hare, 1991) are described in the table footnotes. All predictor variables were coded (without knowledge of any subsequent recidivistic offenses) from the clinical data available on institutional files by trained research assistants.

Phallometric assessment. The phallometric assessment (including stimuli and scoring) procedure and data establishing its discriminative validity have been described in detail elsewhere (Chaplin, Rice, & Harris, 1995; Harris, Rice, Quinsey, & Chaplin, 1996; Harris, Rice, Quinsey, Chaplin, & Earls, 1992). Briefly, child molesters received visual stimuli to assess age and gender preference (Harris et al., 1996) and an aural set to assess interest in coercive sexual activities with children (Quinsey & Chaplin, 1988a). For rapists, the phallometric assessment included descriptions of neutral activities, consenting sex, brutal rapes, and nonsexual violence (Quinsey, Chaplin, & Varney, 1981; Rice, Chaplin, Harris, & Coutts, 1994). In all cases, the penile responses were recorded with a plethysmograph and a mercury-in-silastic strain gauge around the shaft of the offender's penis. Baseline was measured during the first 2 s of each trial, and recording continued for 30 s after stimulus offset. Procedures to inhibit the ability of assessees to dissimulate their responses were also used

Table 1
Sample Characteristics and Performance of Individual Predictor Variables From Actuarial Tools

Variable	Source	Summary[a]	Interrater reliability[b]	Violent[c]	Sexual[d]
Separation from parents under age 16 years (%)	V, S	60	**.78**	**.12**	.09
Elementary school maladjustment score[e]	V, S	2.44 (1.01)	**.96**	**.22**	**.17**
Alcohol abuse history score[f]	V, S	3.26 (2.49)	**.94**	**.20**	**.12**
Never married (%)	V, S	44	**1.0**	**.18**	.05
Nonviolent criminal history score for arrests[g]	V, S	15.5 (26.2)	**.98**	**.23**	**.12**
Violent nonsexual criminal history score for arrests[g]	S	3.86 (8.86)	**1.0**	**.23**	**.26**
Sexual criminal history score for convictions[g]	S, R, 99	4.57 (11.0)	**1.0**	.08	**.14**
Prior admissions to correctional institutions	99	1.53 (1.43)	**.96**	**.22**	**.16**
Met *DSM–III* criteria for personality disorder (%)	V, S	63	**1.0**	**.25**	**.17**
Met *DSM–III* criteria for schizophrenia (%)	V, S	4		−.05	−.02
Any noncontact sex offenses (%)	99	2		.01	.02
Failure on a prior conditional release (%)	V, S	65	**1.0**	**.18**	**.11**
Age at index offense	V, S	30.4 (11.1)	**.97**	**−.28**	**−.20**
Victim injury score for index offense[h]	V, 99	2.21 (1.54)	**.99**	.05	.04
Any female victim (%)	V, S	88	**.88**	−.04	−.03
Any male victim (%)	S, R, 99	12	**.89**	.00	**.10**
Any child (under age 14 years) victim (%)	S	49	**.87**	−.09	.03
Any unrelated victims (%)	R, 99	65	**.87**	**.18**	**.21**
Any stranger victims (%)	R, 99	34		**.17**	**.20**
Phallometric deviance differential[i]	S	1.35 (2.23)		**.14**	.09
Hare PCL–R score[j]	V, S	18.23 (8.51)	**.95**	**.32**	**.17**
Age at risk	99	35.7 (11.6)	**.97**	**−.21**	**−.18**

Note. Significant (α = .05, two-tailed) correlations are shown in bold. V = Violence Risk Appraisal Guide; S = Sex Offender Risk Appraisal Guide; R = Rapid Risk Assessment for Sex Offender Recidivism; 99 = Static-99; *DSM–III* = *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.); PCL–R = Psychopathy Checklist—Revised.

[a] Summary statistics are percentages for dichotomous variables and means accompanied by standard deviations (in parentheses) for continuous variables. [b] Pearson correlations for continuous variables and kappas for dichotomous variables. [c] Point-biserial correlations with violent (including sexual) recidivism. [d] Point-biserial correlations with sexual recidivism only. [e] Rated on a 4-point scale from 1 (*no problems*) to 4 (*severe discipline or attendance problems*). [f] Total score based on 1 point each for parental alcohol problem, teenage alcohol problem, alcohol involved in prior offense, alcohol involved in index offense, and current alcohol problem (fully described in Quinsey, Harris, Rice, & Cormier, 1998). [g] Total score for all qualifying offenses using a modification of the system devised by Akman and Normandeau (1967; fully described in Quinsey, Harris, et al., 1998). [h] Rated on a 7-point scale from 1 (*no injury*) to 7 (*death with mutilation afterwards*). [i] See text in the Method section. [j] Using 20-item version fully described in Hare (1991) scored using documentary information only.

during testing (Harris, Rice, Chaplin, & Quinsey, 1999; Quinsey & Chaplin, 1988b). Phallometric data were obtained for 95 members of the Kingston sample. The stimuli and procedures for those assessments have been described elsewhere (Malcolm, Andrews, & Quinsey, 1993) and were similar to those used at the Penetanguishene (Oak Ridge and community) site. A deviance differential was calculated by subtracting the largest average response to a nondeviant category from the largest average response to a deviant category. A positive score reflected an absolute preference for deviant sexual targets or activities, and the magnitude of the index reflected the difference in standard deviation units. Consistent with procedures established in earlier research cited above, no participants were excluded because of low responding (Harris et al., 1992).

Recidivism. The primary source of the independently coded outcome data was records of charges and convictions of the Fingerprint Service of the Royal Canadian Mounted Police (a national register). Participants were classified as violent recidivists if they incurred a new criminal charge for an offense against persons (e.g., homicide, attempted homicide, assault causing bodily harm, armed robbery, kidnapping) after being released from the study institution. If a violent offense could be ascertained from the criminal record to have been sexually motivated (i.e., sexual assault or sexual interference), the participant was also recorded as a sexual recidivist. Among these sex offenders, many violent reoffenses (some of which were, in reality, sexually motivated) were not coded as sexual recidivism because (due to plea bargaining or lack of evidence or because the victim died and the offense was murder) the word *sexual* did not appear on the police record. Sexual recidivism was a subset, therefore, of violent recidivism. Also included were known violent or sexual offenses that occurred after release from the study institutions, even if formal charges were not made ($n = 6$). Time until failure was calculated from the date the participant was released from the study institution until the date of the last follow-up information or the date of any subsequent violent or sexual offense, whichever came first. Time spent in custody for nonviolent (or nonsexual) offenses was not counted.

Procedure and Analytic Strategy

The coding of all predictor variables was done by three teams of research assistants (one at each site), was based entirely on documentary material compiled before offenders' release, and was conducted independently from the coding of recidivism variables. This was done to prevent contamination of the coding of outcome variables by raters' knowledge of predictors, and vice versa. Actuarial instruments developed in earlier studies (Hanson, 1997; Hanson & Thornton, 2000; Harris et al., 1993; Quinsey, Harris, et al., 1998; Rice & Harris, 1997) were tested in the participant population as a whole, for rapists and child molesters separately, and in each of four subsamples individually. The predictive accuracy was evaluated using ROC statistics (SPSS, Version 9.0).

Interrater reliability was assessed by randomly selecting 10 participants from the two Penetanguishene samples for independent coding of the study variables. Pearson correlation coefficients are shown in Table 1. Interrater reliability was not assessed for the recording of phallometric test results, and for two other variables (noncontact sex offenses and *Diagnostic and Statistical Manual of Mental Disorders* [3rd ed.; *DSM–III*[1]; American Psychiatric Association, 1980] criteria for schizophrenia), there was no variance in the interrater reliability sample and agreement was 100%. Reliability of the actuarial instruments was assessed by comparing scores generated by the two independent codings. Intraclass correlation coefficients were .96 (95% confidence interval [CI] = .84, .99) for the VRAG, .95 (95% CI = .81, .99) for the RRASOR, .87 (95% CI = .57, .97) for the Static-99, and .96 (95% CI = .86, .99) for the SORAG.

Results

Summary statistics characterizing the histories of the offenders, for all study variables, are shown in Table 1, including the bivariate correlations between each independent variable and each of the two outcome variables, violent (including sexual) and sexual recidivism. Table 2 shows mean follow-up times, summary statistics, and initial comparisons for the actuarial instruments. Also shown in Table 2 are the mean proportion of unscorable items for each actuarial instrument at each site. Most commonly, an item was unscorable when the documentary material did not contain enough information to permit coding or the material contained conflicting information. The sites differed in the proportion of items that were coded by the research assistants. For example, missing VRAG items were significantly greater in number for the Penetanguishene community sample than for the other three samples, one-way analysis of variance, $F(3, 392) = 29.08$, $p < .01$. The four instruments were intercorrelated, especially the VRAG and the SORAG ($r = .93$) and the RRASOR and the Static-99 ($r = .87$). The RRASOR was moderately correlated with the VRAG and the SORAG ($r = .31$ and .45, respectively), whereas the Static-99′s correlations with the VRAG and the SORAG were somewhat greater ($r = .49$ and .64, respectively; all ps $< .01$). Figure 1 shows the ROCs for each instrument's prediction of the outcome for which it was developed.

Table 3 shows the comparison of the four actuarial instruments across the subsamples and in the entire sample for both outcome measures (Oak Ridge and community samples were combined for the Penetanguishene sample because they were coded by the same team). As described in the Method section, to evaluate the accuracy of the instruments for the prediction of sexual recidivism, we reclassified offenders who had been subsequently charged with violent reoffenses (no matter how violent) without apparent sexual motivation as nonrecidivists. This meant, for example, that 4 men charged with subsequent homicides (for which the motivation was unclear) were recoded as nonrecidivists. The base rate of sexual recidivism, 26% (104 out of 396), was therefore considerably lower than the base rate of violent recidivism.[2]

Table 4 shows how each instrument performed in predicting each outcome for child molesters and rapists separately (participants who had both adult and child victims are represented in both categories). The base rate of violent recidivism was 35% (60 out of 170) among child molesters and 53% (119 out of 226)

[1] The VRAG and the SORAG use *DSM–III* (American Psychiatric Association, 1980) criteria because that was the current version when they were developed.

[2] It might seem that multilevel analyses would show whether relations between predictors and outcomes vary as a function of various research sites. Although not a focus of the present research, we conducted multivariate analyses of variance to evaluate this possibility. The analyses yielded two significant main effects—that is, some sites yielded better prediction than others and violent recidivism was better predicted. However, no statistically significant interactions resulted. Thus, this multilevel approach revealed no findings that were not evident in the simpler comparisons reported here.

Table 2
Performance of the Actuarial Instruments Predicting Violent Recidivism in Four Samples

Measure	Oak Ridge	Community	Kingston	Pacific	Total
N	118	87	96	95	396
Violent recidivism rate	.52	.22	.60	.49	.48
Months of opportunity	51.8 (58.4)	42.2 (24.9)	58.9 (34.1)	95.5 (48.8)	61.5 (48.1)
VRAG					
M (*SD*)	11.3 (11.2)	1.39 (8.76)	6.44 (11.4)	6.77 (9.50)	7.28 (11.0)
% missing items[a]	8	19	7	8	10
r with violent recidivism	.40	.37	.36	.32	.40
SORAG					
M (*SD*)	15.2 (13.3)	2.15 (11.4)	9.03 (14.2)	9.38 (12.3)	9.99 (10.8)
% missing items[a]	8	15	7	14	11
r with violent recidivism	.34	.37	.36	.31	.38
RRASOR					
M (*SD*)	2.33 (1.56)	1.54 (1.63)	1.48 (1.56)	2.09 (1.70)	1.89 (1.65)
% missing items[a]	7	2	8	4	6
r with violent recidivism	.08*	.12*	.14*	.08*	.11
Static-99					
M (*SD*)	4.32 (2.24)	2.38 (2.27)	2.84 (2.23)	3.91 (2.40)	3.44 (2.41)
% missing items[a]	9	10	12	10	10
r with violent recidivism	.13*	.18	.25	.17	.21

Note. VRAG = Violence Risk Appraisal Guide; SORAG = Sex Offender Risk Appraisal Guide; RRASOR = Rapid Risk Assessment for Sex Offender Recidivism.
[a] Mean percentage of items that could not be scored because of missing data.
* $p > .05$, one-tailed.

among rapists, $\chi^2(1, N = 396) = 11.41$, $p < .01$. Meanwhile, the groups were more similar in their rates of sexual recidivism: 21% (35 out of 170) and 29% (66 out of 226) for child molesters and rapists, respectively, $\chi^2(1, N = 396) = 3.22$, $p < .10$. Significantly better (on the basis of CIs) prediction of violent recidivism was generally achieved by the VRAG and the SORAG for both offender types.[3]

It was also of interest to examine the accuracy of the probability estimates given for the individual VRAG and SORAG bins in the previously published norms (Quinsey, Harris, et al., 1998). The shortest mean follow-up period used in the construction samples was approximately 7 years, whereas the mean opportunity for the present participants was considerably shorter, 61.4 months ($SD = 48.5$) or approximately 5 years. Nevertheless, as shown in Table 5, the rates of violent recidivism from the present results were usually slightly higher than expected on the basis of the norms. Chi-square goodness-of-fit tests (in which the expected values were given by the norms and the observed values came from the present results) indicated only marginally significant differences for both the VRAG and the SORAG, $\chi^2(7, N = 396) = 13.14$, $p < .10$, and $\chi^2(8, N = 396) = 14.60$, $p < .10$, respectively. The VRAG and the SORAG yielded almost identical and small absolute mean differences between the obtained and expected rates, .088 and .089, respectively. These differences mean that, on average, the probabilities from the normative bins were less than .09 different from the obtained probabilities. In addition, the mean base rate of violent recidivism in this study (48%) was similar to that reported in earlier studies (e.g., 40%; Quinsey, Rice, & Harris, 1995) over similar follow-up durations. Table 6 shows the same comparison for the RRASOR and the Static-99 for sexual recidivism for the 5-year follow-up period specified in the available norms (Hanson, 1997; Hanson & Thornton, 2000). For both, chi-square goodness-of-fit analyses indicated significant deviation from published norms: $\chi^2(6, N = 396) = 290.45$, $p < .01$, and $\chi^2(5, N = 396) = 101.29$, $p < .01$, for the Static-99 and the RRASOR, respectively. Overall, the present replication of those actuarial instruments appears to exhibit a trend toward regression to the mean (nonsignificantly for the VRAG and the SORAG). Results reported elsewhere (Harris & Rice, 2003) suggest that this trend was due to the effects of missing items.

Some investigators adopt a constant follow-up period (e.g., Epperson, Kaul, & Hesselton, 1998) by eliminating nonrecidivists with less than a set duration of opportunity and considering those who fail after that set duration as successes. We examined the effect of such a procedural decision in the present data by adopting a fixed follow-up period of 36 months. Thus, any participant with less opportunity who had not recidivated was dropped, and any participant who recidivated later than 36 months after release was reclassified as a nonrecidivist. This change raised the accuracy of the VRAG in predicting violent

[3] Throughout this article, inferences about statistically significant differences between ROC areas are based on 95% CIs derived from maximum-likelihood estimates of the ROC functions. Thus, an ROC area obtained for a particular test that lay outside the 95% CI for another test was inferred to be reliably different. A different, slightly less conservative, method using *z* scores was given by Hanley and McNeil (1983), which was also applied to all relevant pairwise comparisons in the present data. The only test resulting in a different conclusion than that based on CIs was the overall comparison of RRASOR and Static-99 predicting sexual recidivism (lower right corner of Table 3); the method of Hanley and McNeil indicated greater accuracy for the Static-99 ($z = 2.0$, $p < .05$).

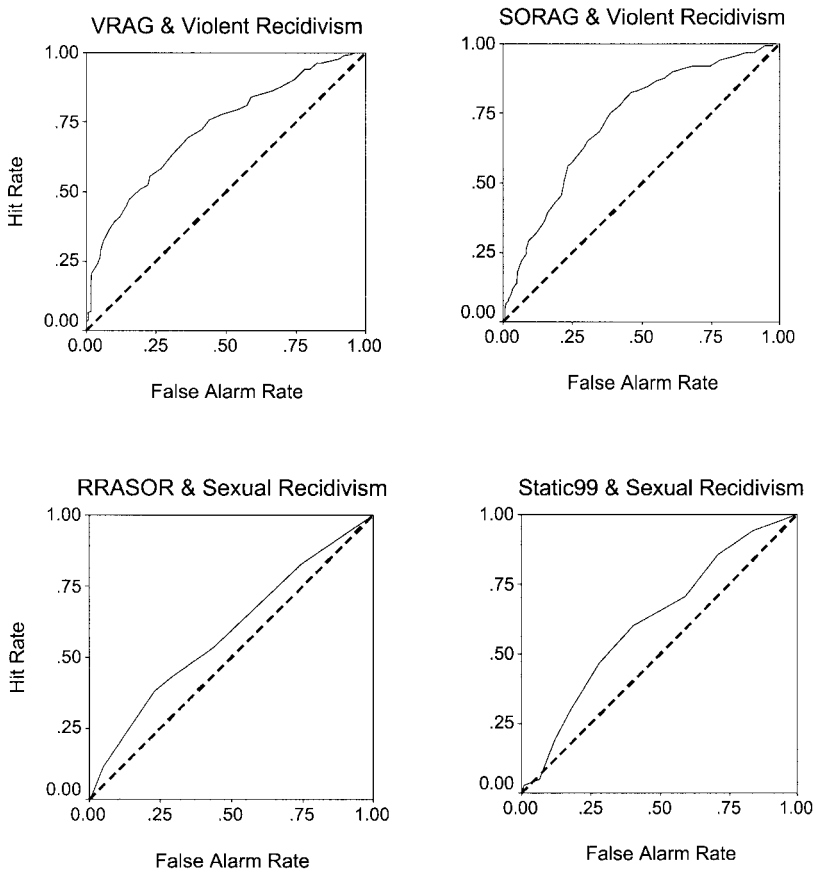


Figure 1. Receiver operating characteristics (ROCs) for four actuarial instruments in predicting the outcome for which each was developed. VRAG = Violence Risk Appraisal Guide; SORAG = Sex Offender Risk Appraisal Guide; RRASOR = Rapid Risk Assessment for Sex Offender Recidivism.

recidivism from an ROC area of .73 (Table 3) to .75 (*r* increased from .40 to .43; Cohen's *d* increased from .87 to 1.00), whereas the base rate went from .48 to .33 (115 out of 348). The effect of the constant follow-up period was greater when applied to the Static-99′s ability to predict sexual recidivism. ROC area increased from .62 (Table 3) to .70 (*r* increased from .17 to .31; Cohen's *d* increased from .40 to .75), whereas the base rate went from .27 to .24 (74 out of 307). Converting to constant follow-up raised ROC areas for all instruments and both outcomes.

Next, we note the contribution of complete data for prediction. The ROC area for prediction of violent recidivism with the VRAG was .73, as shown in Table 2; this increased to .80 when only participants (n = 46) who had no missing data[4] were considered. When participants with no missing data were considered for an exact follow-up of 36 months, the accuracy of prediction was optimized. For the VRAG's ability to predict violent recidivism (base rate = .48), ROC area was .84 (SE = .06), and the correlation was .55 (Cohen's d = 1.21), comprising a statistically significant ($p < .05$) improvement (based on CIs) over the performance of the VRAG overall. A similar pattern of results was obtained for the other instruments and outcomes.

Kaplan–Meier Survival Curves (Luke & Homan, 1998) for violent recidivism for the four samples are shown in Figure 2A, illustrating the clear differences in risk exhibited by the groups. To illustrate the size of various effects, we computed Kaplan–Meier survival curves for PCL–R scores (≥25 vs. <25; Figure 2B) and phallometrically determined sexual deviance (absolute overall preference vs. no preference for deviant stimuli; Figure 2C). These survival data describe sex offenders' long-term risk of violent

[4] For the purposes of this analysis, having no missing data meant that all 20 items on the PCL–R could be scored and each of the other 11 VRAG items was available. In addition, for this analysis, complete data on the net elapsed time until recidivism or the study end date were also required. The various measures used in this study differ somewhat in how missing data are handled. The PCL–R manual (Hare, 1991) allows as many as 5 unscorable items, and prorating is used to arrive at a final score. The VRAG and SORAG manual (Quinsey, Harris, et al., 1998) recommends no more than 4 missing items, and the scoring system (using positive and negative weights) simply omits unscoreable items. The available scoring instructions for the RRASOR and the Static-99 (Hanson, 1997; Hanson & Thornton, 1999) contain no explicit instructions about missing items; for all analyses reported here, unscorable items were omitted.

Table 3
Comparison of Actuarial Prediction Tools Using Area Under the Receiver Operating Characteristic (ROC) by Outcome and Site

	Penetanguishene		Kingston		Pacific		All	
Instrument	ROC (*SE*)	CI	ROC (*SE*)	CI	ROC (*SE*)	CI	ROC (*SE*)	CI
				Violent recidivism				
VRAG	.77 (.03)	.71, .84	.70 (.05)	.60, .81	.70 (.06)	.59, .80	.73 (.03)	.68, .78
SORAG	.77 (.03)	.70, .83	.71 (.05)	.61, .82	.69 (.06)	.58, .79	.73 (.03)	.68, .78
RRASOR	.59 (.04)	.51, .67	.60 (.06)	.48, .72	.55 (.06)	.43, .67	.56 (.03)	.51, .62
Static-99	.66 (.04)	.59, .74	.67 (.06)	.55, .78	.60 (.06)	.48, .71	.63 (.03)	.57, .68
				Sexual recidivism				
VRAG	.72 (.04)	.64, .80	.61 (.06)	.51, .73	.61 (.06)	.48, .73	.65 (.03)	.59, .71
SORAG	.71 (.04)	.63, .79	.62 (.06)	.51, .74	.59 (.06)	.46, .71	.66 (.03)	.60, .71
RRASOR	.63 (.05)	.54, .73	.61 (.06)	.50, .73	.52 (.07)	.39, .65	.59 (.03)	.52, .65
Static-99	.67 (.05)	.58, .76	.63 (.06)	.52, .75	.54 (.07)	.41, .66	.62 (.03)	.56, .68

Note. ROC represents the maximum-likelihood estimates of area under the ROC. CI = 95% confidence interval; VRAG = Violence Risk Appraisal Guide; SORAG = Sex Offender Risk Appraisal Guide; RRASOR = Rapid Risk Assessment for Sex Offender Recidivism.

recidivism[5] and demonstrate that identifiable subgroups of sex offenders (i.e., psychopathic offenders, sexually deviant offenders) exhibit long-term risk of violence at rates so high as to have social and criminal justice policy implications. A related concern pertains to the prediction of recidivism for very short or long follow-up times. Litwack (2001), for example, asserted that the task of short-term prediction is quite different from long-term prediction. Using data from all the present participants, we simulated various minimum follow-up times by eliminating nonrecidivating participants without minimum opportunity (shown in Figure 3). Thus, Figure 3 illustrates the accuracy (expressed as ROC area) with which the VRAG would have predicted violent recidivism if all participants had been followed for 1 year or 2 years and so on up to at least 13 years. Clearly, although the base rate of violent recidivism approaches unity, there was no statistically significant effect on VRAG accuracy.

Table 1 and Figure 2 show that both phallometric deviance differential and PCL–R scores were correlated with violent and sexual recidivism. We also tested whether their statistical interaction added to predictive accuracy. First, binary logistic regression indicated a statistically significant effect for the interaction in the prediction of violent recidivism, model $\chi^2(1, N = 155) = 13.82$, $p < .01$, Wald's statistic = 10.18, $p < .01$, and a marginal effect for the prediction of sexual recidivism, model $\chi^2(1, N = 155) = 3.57$, $p < .06$, Wald's statistic = 3.20, $p < .10$. A more powerful test of the interaction's effects was afforded by Cox regression survival analysis, incorporating the time until recidivism: For violent recidivism, overall model $\chi^2(1, N = 147) = 7.67$, $p < .01$, Wald's statistic = 7.37, $p < .01$; and for sexual recidivism, overall model $\chi^2(1, N = 147) = 5.72$, $p < .05$, Wald's statistic = 5.39, $p < .05$. Survival functions, based on median splits for the two predictor variables (PCL–R *Mdn* = 17.5; phallometric deviance differential *Mdn* = 1.13), for both outcomes are shown in Figure 4. Although statistical analyses indicated significant interactions, Figures 2 and 4 also indicate main effects for PCL–R scores in predicting both outcomes, Wald's statistic = 33.42, $p < .01$, and Wald's statistic = 7.08, $p < .01$, respectively. Survival analyses did not yield significant main effects for phallometric deviance; however, as shown in Table 1, it was significantly correlated with dichotomous violent recidivism.

Of course, dichotomous recidivism was the outcome each of the present actuarial instruments was designed to predict. Nevertheless, as implied by the survival analyses, each also predicted the speed of recidivism, Pearson $r = .33$ ($n = 162$), $p < .01$, for the VRAG and the SORAG; $r = .23$ ($n = 90$), $p < .05$, for the RRASOR; and $r = .36$ ($n = 90$), $p < .01$, for the Static-99. On a 20-point ordinal scale from property offenses to first-degree murder, the VRAG and the SORAG predicted the severity of the outcome ($r = .21$, $p < .01$, and $r = .18$, $p < .01$, respectively; $ns = 240$). Finally, both the VRAG and the SORAG also predicted the severity of injury to victims (on a 7-point scale from 1 = *none* to 7 = *death and mutilation*) in the recidivistic offenses ($r = .35$, $p < .01$, and $r = .30$, $p < .01$, respectively; $ns = 71$). The

[5] Survival analysis is a set of statistical procedures used to discover relationships between variables and outcome events that incorporate the passage of time until the event occurs. One of the most powerful features of survival analysis is the ability to deal with censoring—the fact that participants vary in how long they are followed. Survival analyses are particularly suited to studies of recidivism because the length of time a released offender remains free of criminal behavior often has as much practical and theoretical significance as whether recidivism occurs at all and because censoring is often related to variables of interest—for example, offenders with extensive criminal histories might experience shorter opportunity because conditional release is denied. Survival techniques include inferential tests for group differences, the ability to test multiple predictors simultaneously for independent contributions, comparison of effect sizes, and tests of the effects of variables that depend on the passage of time (Luke & Homan, 1998).

Table 4
Comparison of Actuarial Prediction Tools by Outcome and Offender Type

Instrument	Violent recidivism				Sexual recidivism			
	Recidivists	Nonrecidivists	ROC	95% CI	Recidivists	Nonrecidivists	ROC	95% CI
				Child molesters				
VRAG	8.17 (11.3)	0.43 (8.22)	.70 (.05)	.61, .78	9.31 (11.3)	1.56 (9.15)	.70 (.06)	.59, .81
SORAG	9.13 (12.5)	−0.33 (9.59)	.72 (.04)	.63, .80	9.83 (12.5)	1.24 (10.7)	.70 (.05)	.59, .80
RRASOR	2.08 (1.88)	1.34 (1.63)	.61 (.05)	.52, .70	2.20 (1.98)	1.44 (1.66)	.61 (.06)	.50, .72
Static-99	3.25 (2.45)	2.09 (2.19)	.64 (.05)	.55, .73	3.49 (2.50)	2.24 (2.24)	.65 (.05)	.54, .75
				Rapists				
VRAG	13.5 (10.1)	4.78 (9.85)	.73 (.03)	.66, .79	13.6 (10.3)	7.64 (10.6)	.64 (.04)	.56, .72
SORAG	18.2 (11.7)	8.75 (13.6)	.70 (.04)	.63, .76	18.1 (11.8)	12.0 (13.7)	.62 (.04)	.54, .69
RRASOR	2.25 (1.58)	2.22 (1.52)	.50 (.04)	.42, .57	2.50 (1.68)	2.13 (1.48)	.56 (.04)	.47, .64
Static-99	4.54 (2.15)	3.85 (2.23)	.58 (.04)	.51, .66	4.70 (2.11)	4.01 (2.22)	.59 (.04)	.51, .66

Note. For recidivists and nonrecidivists, the values are means (and standard deviations). For ROC, the values are area under the maximum-likelihood receiver operating characteristic (with standard errors in parentheses). CI = confidence interval; VRAG = Violence Risk Appraisal Guide; SORAG = Sex Offender Risk Appraisal Guide; RRASOR = Rapid Risk Assessment for Sex Offender Recidivism.

RRASOR and the Static-99 were unrelated to these latter two indices of the severity of recidivism.

Discussion

Among four samples of sex offenders from three different sites, all four actuarial instruments examined in the present study reliably predicted both violent (including sexual) recidivism and recidivism known to be sexually motivated. ROC areas were consistently higher for the VRAG and the SORAG than for the RRASOR or the Static-99 for both outcomes, although the differences among the instruments were not always statistically significant for sexual recidivism. Using conventional standards for classifying effect sizes (Cohen, 1992), Table 3 shows that over all sites, effect sizes for the VRAG and the SORAG were large for violent recidivism and moderate for sexual recidivism. Effect sizes for the RRASOR and the Static-99 were small to moderate for both outcomes. For the site (Oak Ridge) with the least missing information and known high reliability of scoring, effect sizes for the VRAG and the SORAG were large for both outcomes, moderate for both outcomes for the Static-99, and small for the RRASOR. Both the VRAG and the SORAG predicted imminence and severity of recidivism. The relative accuracies of the performance of the four actuarial scales is attested to by their relatively consistent ROC areas over sites that vary geographically, in whether sex offenders were psychiatric patients or penitentiary inmates, in the base rate of recidivism, and in whether offenders were institutionalized or living in the community at the time of assessment. This consistent ranking occurred using both violent (including sexual) recidivism and sexual recidivism known to be sexual as the outcome measure.

Overall, there was a tendency toward higher effect sizes for the prediction of what we defined as "violent," rather than "sexual," recidivism. This effect was entirely attributable to the VRAG and the SORAG. For the RRASOR and the Static-99, the predictive abilities for violent and sexual recidivism were virtually identical, even though both were developed to predict sexual recidivism specifically. The results lead us to wonder whether what we defined as violent recidivism may be a more valid (i.e., less

Table 5
Observed Rates of Violent (Including Sexual) Recidivism for Each of Nine VRAG and SORAG Categories or Bins (After Mean 5 Years of Opportunity) and Corresponding Expected Values for 7 Years Based on Norms Provided in Quinsey, Harris, Rice, and Cormier (1998)

Category or bin	VRAG		SORAG	
	Expected	Observed	Expected	Observed
1	.00		.07	.19
2	.08	.00	.15	.18
3	.12	.20	.23	.29
4	.17	.31	.39	.50
5	.35	.39	.45	.55
6	.44	.51	.58	.63
7	.55	.65	.58	.63
8	.76	.84	.75	.71
9	1.0	.89	1.0	.76

Note. VRAG = Violence Risk Appraisal Guide; SORAG = Sex Offender Risk Appraisal Guide.

Table 6
Observed (O) and Expected (E) Recidivism Rates Over 5 Years for the Static-99 and the Rapid Risk Assessment for Sex Offender Recidivism (RRASOR)

Score	Static-99		RRASOR	
	E (sexual)	O (sexual)	E (sexual)	O (sexual)
0	.05	.11	.04	.20
1	.06	.19	.08	.26
2	.09	.31	.14	.20
3	.12	.17	.25	.23
4	.26	.28	.33	.35
5	.33	.38	.50	.46
6+	.39	.37		

"noisy") measure of subsequent violent (i.e., hands-on) sex offenses than is sexual recidivism. As suggested elsewhere (Rice & Harris, 1999), there is reason to believe that, because of such factors as plea bargaining, offenses that are truly sexually motivated result in charges or convictions that appear on criminal records as nonsexual violent offenses. Perhaps even if the only outcome of interest for sex offenders were the commission of future sex offenses, our measure of arrests or convictions for violent offenses is a more valid measure than sexual recidivism as usually operationalized in studies of sex offender recidivism. We hope to investigate this in further empirical research.

The observation that predictive accuracy as measured by the ROC area, and, therefore, the numerically equivalent common language effect size (McGraw & Wong, 1992), was higher with a fixed rather than a variable follow-up period means that these statistics are not strictly comparable between follow-up studies that use the two different procedures. Moreover, this finding implies a negative mathematical relationship between the variance in follow-up time over participants and measures of predictive accuracy. Presumably, a correction factor could be developed to allow investigators to estimate the accuracy that would be obtained in a follow-up of fixed length from the variance in follow-up time in a given study. Another important factor affecting the size of prediction accuracy was the availability of complete information. Indeed, the accuracy of the VRAG, for example, in predicting violent recidivism increased from .73 to .80 (ROC area) when we included only participants with complete information and to .84 when there was both complete information and an exact follow-up time.

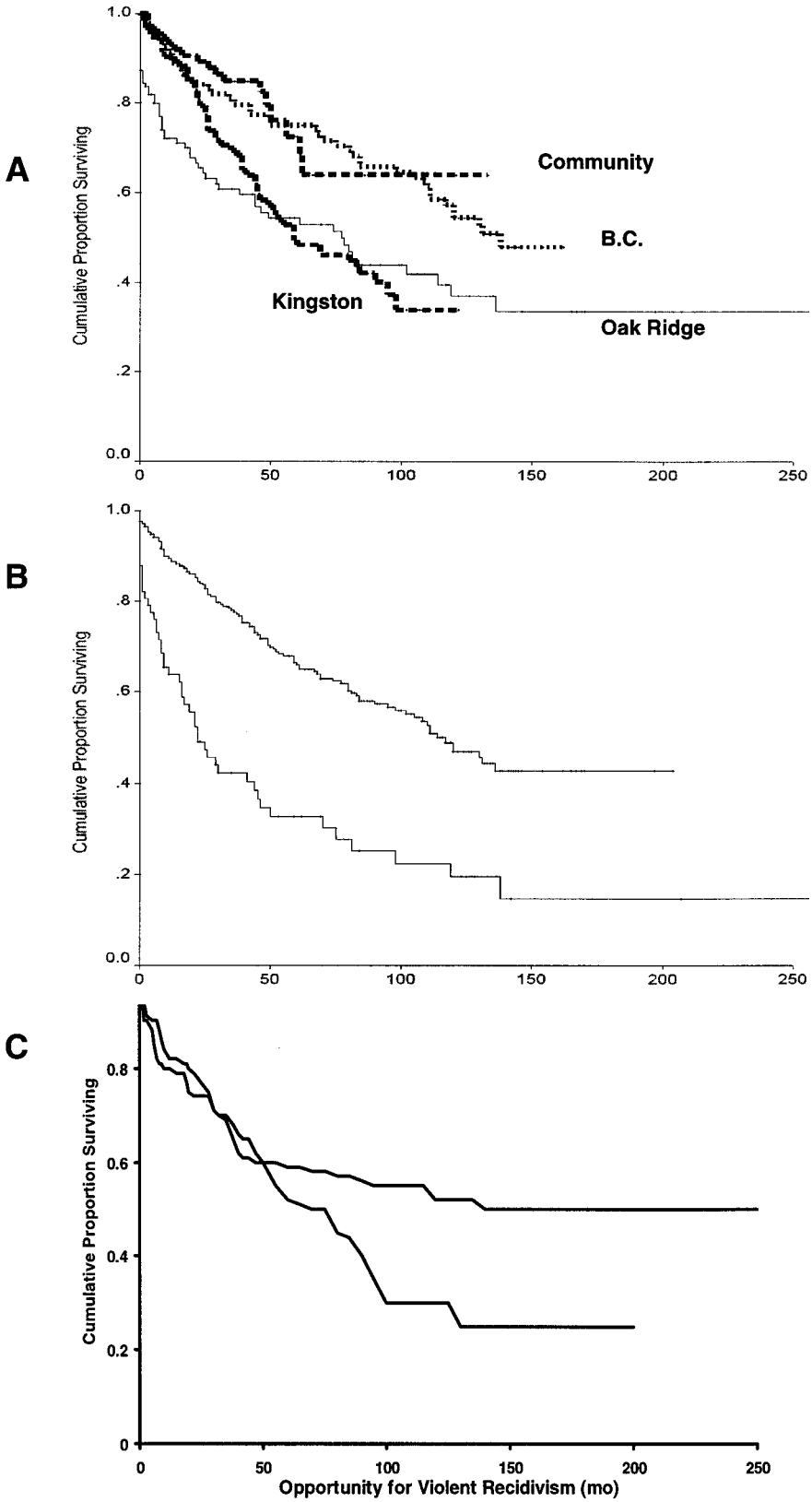
The findings that the accuracy of actuarial instruments varied according to the variability of the follow-up time and the amount of missing information offer possible explanations for the findings of apparently different predictive accuracies of some of the instruments across studies. For example, some studies reporting lower predictive accuracies for the VRAG and the SORAG have omitted or altered several items (e.g., Grann, Belfrage, & Tengström, 2000; Nunes et al., 2002; Sjöstedt & Langström, 2002). In addition, there is variability in the quantity and quality of documentary information available to researchers studying the recidivism of sex offenders. Future studies using actuarial instruments should ensure that the variability of the follow-up times is reported and that items omitted or approximated are fully described.

One important finding in the present study was the interaction between psychopathy and phallometrically measured sexual deviance in predicting recidivism. This interaction was found by Rice and Harris (1997) as well as in a follow-up study of juvenile sexual offenders (Gretton et al., 2001). Because of the robustness of this interaction and its prognostic significance, its inclusion in the next generation of actuarial instruments for sex offenders should increase predictive accuracy further. It is noteworthy that this interaction was observed not only in sexual recidivism, for which it would be expected, but also in violent (including sexual) recidivism. Our interpretation of this finding is that, as noted earlier, a substantial proportion of sex offenders' reoffenses that are coded as nonsexual but violent are in fact sexually motivated. This interpretation could also explain why phallometric results were better predictors of violent than sexual recidivism in the present study.

Results from this and other studies clearly support the use of actuarial instruments for determining the dangerousness of adjudicated sex offenders. Actuarial information on dangerousness can be used in at least three ways. The first involves the characterization of a given population. For example, managers of forensic units might be interested in finding out the overall level of risk of groups of clients in order to adjust security practices and interventions accordingly (Rice & Harris, 2003). The second way to use actuarial information is the ranking of individual sex offenders according to risk. Ranking would be useful, for example, in determining levels of security (the most dangerous offenders are placed in the most secure settings), need for specialized interventions, and required intensity of supervision. Ranking is perhaps the safest way to use actuarial information because although the absolute rate of recidivism may vary across time and place, an offender's rank order is likely to be very stable. Third, results from the present study support the use of the VRAG and the SORAG to estimate the probability of recidivism over a given period of opportunity with rapists and child molesters because the observed probabilities were quite close to the normative (Quinsey, Harris, et al., 1998) probabilities.

The first use for actuarials in the previous paragraph focuses on group-level decisions, whereas the latter two uses focus more on the individual level. This distinction between group- and individual-level analysis has a large impact on one's impression of the predictive value of actuarial tools. The situation is somewhat analogous to predicting survival among cancer patients—survival analyses can give a good account of subgroups of patients defined by particular risk factors or treatment approaches (Luke & Homan, 1998) but cannot say precisely what will be the fate of any individual patient. Similarly, the close correspondence between expected and observed recidivism rates for groups defined by scores on the VRAG, for example (shown on the left side of Table 5), yields impressive correspondence at the group level, a perfect rank-order correlation (Pearson and intraclass rs = .96). However, at the individual, case-by-case level, the actuarial tools evaluated in the present study were not perfect. The present case-by-case sensitivity–specificity trade-off captured by the ROC area of .84 under fairly optimal conditions still leaves room for improving predictive accuracy by such means as incorporating better risk factors, including interactions among risk factors, and reducing error in the measurement of outcome. Nevertheless, as was the case in the development sample for the VRAG and the SORAG, the standard error of measurement achieved in the present study (e.g., 2.24 for the VRAG) indicates that the true risk estimates for a particular individual is highly unlikely to vary by more than one category or bin from his obtained score.

All of the actuarial instruments evaluated in the present study use static items that cannot reflect fluctuations in risk. Many of the present participants (36%) completed some form of treatment according to information recorded in their files sometime after the material we used to score the instruments evaluated here. Might prediction have been improved by adding knowledge about treatment to the present actuarial instruments? For each instrument and both outcomes, multivariate analyses indicated that treatment made a nonsignificant positive contribution after actuarial scores were incorporated. Treatment was nonsignificantly associated with increased recidivism. This result is consistent with our review of the scientific literature on sex offender therapy (Rice & Harris, 2003): Until effective treatments for sex offenders are developed, risk assessment is unimpaired by excluding consideration of participation in or response to therapy.


A
Cumulative Proportion Surviving
1.0
.8
.6
.4
.2
0.0
0
50
100
150
200
250
Community
B.C.
Kingston
Oak Ridge
B
Cumulative Proportion Surviving
1.0
.8
.6
.4
.2
0.0
0
50
100
150
200
250
C
Cumulative Proportion Surviving
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
Opportunity for Violent Recidivism (mo)

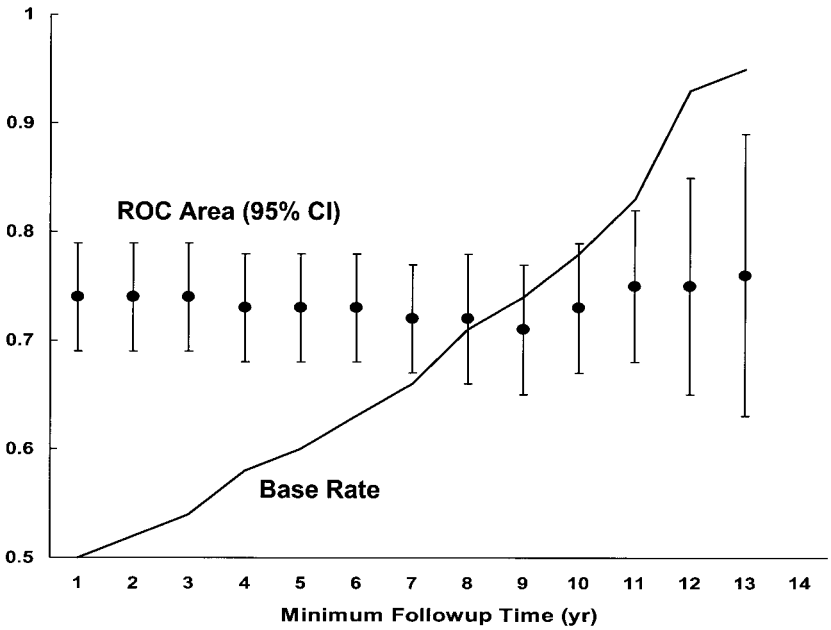


Figure 3. Base rate and receiver operating characteristic (ROC) area with 95% confidence intervals (CI) for Violence Risk Appraisal Guide score in predicting violent recidivism as a function of minimum follow-up time. yr = years.

As described in an earlier article (Rice & Harris, 1995), ROC statistics allow one to select a cutpoint for a particular purpose, such as bail, parole, conditional release, and so forth, according to a priori decisions about the relative costs of false positives and false negatives. Although there are demonstrations of the use of ROC statistics in ascertaining cutoff scores once relative costs have been set (e.g., Rice & Harris, 1995), it is clear that there is no consensus about relative costs in the first place (cf. Mathiesen, 1998; Webster, 1998). Although there has been pilot work (Mossman & Hart, 1993), an important research goal is now the development of an empirically based balance of costs, perhaps using population surveys of the utilities associated with preventing various crimes (e.g., Quinsey, Lalumière, Querée, & McNaughton, 1999).

Although among child molesters the rate of violent (including sexual) recidivism was higher than the rate of recidivism known to be sexually motivated, both outcomes were equally well predicted by each of the four actuarial instruments evaluated. Among rapists, however, there was a greater difference in the base rate of the two outcomes, and the VRAG–SORAG tools predicted violent (including sexual) recidivism significantly better than recidivism known to be sexually motivated. It is interesting to note that rapists were more likely to have psychopathy (PCL–R $M = 19$, $SD = 7.8$; 95% CI = ± 2.00) than were child molesters (PCL–R $M = 15$, $SD = 6.7$; 95% CI = ± 1.00). Perhaps psychopathic individuals were more able to obscure sexual motives of the recidivistic offenses for which they were apprehended.

The statistically significant interaction of psychopathy and sexual deviance in the prediction of both violent and sexual recidivism in sex offenders, the superior performance of prediction tools that included measures of psychopathy or sexual deviance, and the differences between rapists and child molesters in both the base rates and prediction of violent versus sexual recidivism all attest to the centrality of both sexual deviance and psychopathy in any explanation of these two forms of sexual aggression (Lalumière, Quinsey, Harris, & Rice, in press). Although relevant to both, we hypothesize that sexual deviance is the larger contributor to child molestation and that psychopathy is the greater contributor to rape (Lalumière et al., in press). The finding about psychopathy among child molesters and rapists can be stated another way: Among sex offenders, those who were psychopathic (e.g., PCL–R ≥ 30) were more likely to have included women among their victims than those who were not psychopathic (67% vs. 41%), $\chi^2(1, N = 396) = 6.66$, $p < .05$. This is consistent with a view of psychopathy as a life strategy that has been reproductively viable through human evolutionary history (Harris, Rice, & Lalumière, 2001; Harris, Skilling, & Rice, 2001; Lalumière, Harris, & Rice, 2001; Quinsey & Lalumière, 1995). We are testing this hypothesis with further analyses of the present data.

Figure 2 (opposite). Kaplan–Meier survival curves for violent recidivism for (A) the four subgroups; (B) those participants scoring 25 or higher (lower function) on the Psychopathy Checklist—Revised compared with those scoring lower than 25 (upper function); and (C) those participants who, in phallometric testing, exhibited an absolute preference for a deviant stimulus category (i.e., children, rape, nonsexual violence) in at least one test (lower function) compared with those participants who did not (upper function). B.C. = British Columbia; mo = months.

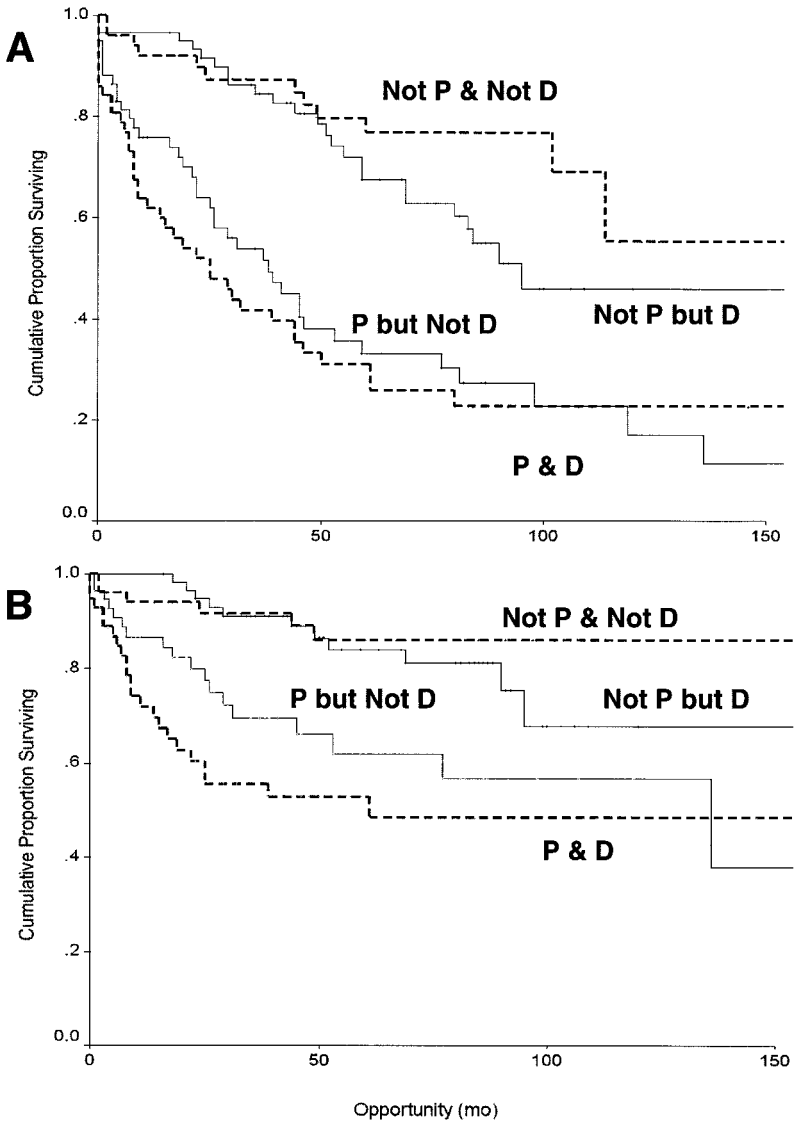


Figure 4. Kaplan–Meier survival curves showing the interaction of psychopathy (P) and sexual deviance (D) on violent recidivism (A) and sexual recidivism (B). Median splits were performed on Psychopathy Checklist—Revised (PCL–R) score and phallometric deviance differential. mo = months.

References

Akman, D. D., & Normandeau, A. (1967). The measurement of crime and delinquency in Canada: A replication study. *British Journal of Criminology, 7,* 129–149.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior, 28,* 490–521.

Bélanger, N., & Earls, C. (1996). Sex offender recidivism prediction. *Forum on Correctional Research, 8,* 22–24.

Chaplin, T. C., Rice, M. E., & Harris, G. T. (1995). Salient victim suffering and the sexual response of child molesters. *Journal of Consulting and Clinical Psychology, 63,* 249–255.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Dempster, R. J., Hart, S. D., & Boer, D. P. (2001). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments.* Manuscript in preparation.

Epperson, D. L., Kaul, J. D., & Hesselton, D. (1998, October). *Final report on the development of the Minnesota Sex Offender Screening Tool—Revised (MnSOST–R).* Paper presented at the 15th Annual Research and Treatment Convention of the Association for the Treatment of Sexual Abusers, Vancouver, British Columbia, Canada.

Glover, A. J. J., Nicholson, D. E., Hemmati, T., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high risk federal offenders. *Criminal Justice and Behavior, 29,* 235–249.

Grann, M., Belfrage, H., & Tengström, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27,* 97–114.

Gretton, H. M., McBride, M., Hare, R. D., O'Shaughnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior, 28,* 427–449.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839–843.

Hanson, R. K. (1997). *The development of a brief actuarial scale for sexual offense recidivism.* Ottawa, Ontario, Canada: Department of the Solicitor General.

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66,* 348–362.

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene: Dynamic predictors of sexual assault recidivism. *Criminal Justice and Behavior, 27,* 6–35.

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders.* Ottawa, Ontario, Canada: Department of the Solicitor General.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24,* 119–136.

Hare, R. D. (1991). *The Hare Psychopathy Checklist—Revised: Manual.* Toronto, Ontario, Canada: Multi-Health Systems.

Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. In R. A. Prentky, E. Janus, & M. C. Seto (Eds.), *Understanding and managing sexually coercive behavior* (pp. 198–210). New York: Annals of the New York Academy of Sciences.

Harris, G. T., Rice, M. E., Chaplin, T. C., & Quinsey, V. L. (1999). Dissimulation in phallometric testing of rapists' sexual preferences. *Archives of Sexual Behavior, 28,* 223–232.

Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26,* 377–394.

Harris, G. T., Rice, M. E., & Lalumière, M. L. (2001). Criminal violence: The roles of psychopathy, neurodevelopmental insults, and antisocial parenting. *Criminal Justice and Behavior, 28,* 402–406.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315–335.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy, 34,* 389–394.

Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment. *Psychological Assessment, 4,* 502–511.

Harris, G. T., Skilling, T. A., & Rice, M. E. (2001). The construct of psychopathy. In M. Tonry (Ed.), *Crime and justice: An annual review of research* (pp. 197–264). Chicago: University of Chicago Press.

Lalumière, M. L., Harris, G. T., & Rice, M. E. (2001). Psychopathy and developmental instability. *Evolution and Human Behavior, 22,* 75–92.

Lalumière, M. L., Quinsey, V. L., Harris, G. T., & Rice, M. E. (in press). *The nature of rape.* Washington, DC: American Psychological Association.

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7,* 409–443.

Luke, D. A., & Homan, S. M. (1998). Time and change: Using survival analysis in clinical assessment and treatment evaluation. *Psychological Assessment, 10,* 360–378.

Malcolm, P. B., Andrews, D. A., & Quinsey, V. L. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence, 8,* 486–501.

Mathiesen, T. (1998). Selective incapacitation revisited. *Law and Human Behavior, 22,* 455–470.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111,* 361–365.

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Clark Robbins, P., Mulvey, E. P., et al. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence.* New York: Oxford University Press.

Mossman, D., & Hart, K. J. (1993). How bad is civil commitment? A study of attitudes toward violence and involuntary hospitalization. *Bulletin of the American Academy of Psychiatry and the Law, 21,* 181–194.

Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research and Treatment, 14,* 253–269.

Quinsey, V. L., Book, A., & Skilling, T. A. (2003). *A follow-up of deinstitutionalized developmentally handicapped men with histories of antisocial behavior.* Manuscript submitted for publication.

Quinsey, V. L., & Chaplin, T. C. (1988a). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence, 3,* 259–274.

Quinsey, V. L., & Chaplin, T. C. (1988b). Preventing faking in phallometric assessments of sexual preference. In R. A. Prentky & V. L. Quinsey (Eds.), *Human sexual aggression: Current perspectives* (pp. 49–58). New York: New York Academy of Sciences.

Quinsey, V. L., Chaplin, T. C., & Varney, G. (1981). A comparison of rapists' and non-sex offenders' sexual preferences for mutually consenting sex, rape, and physical abuse of women. *Behavioral Assessment, 3,* 127–135.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association.

Quinsey, V. L., Khanna, A., & Malcolm, P. B. (1998). A retrospective evaluation of the regional treatment centre sex offender treatment program. *Journal of Interpersonal Violence, 13,* 621–644.

Quinsey, V. L., & Lalumière, M. L. (1995). Evolutionary perspectives on sexual offending. *Sexual Abuse: A Journal of Research and Treatment, 7,* 301–315.

Quinsey, V. L., Lalumière, M. L., Querée, M., & McNaughton, J. K. (1999). Perceived crime severity and biological kinship. *Human Nature, 10,* 399–414.

Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence,* 10, 85–105.

Rice, M. E., Chaplin, T., Harris, G. T., & Coutts, J. (1994). Empathy for the victim and sexual arousal among rapists and nonrapists. *Journal of Interpersonal Violence, 9,* 435–449.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63,* 737–748.

Rice, M. E., & Harris, G. T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior, 21,* 231–241.

Rice, M. E., & Harris, G. T. (1999). Sexual aggressors. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence: The law and science of expert testimony* (pp. 89–121). St. Paul, MN: West.

Rice, M. E., & Harris, G. T. (2002). Men who molest their sexually immature daughters: Is a special explanation required? *Journal of Abnormal Psychology, 111,* 329–339.

Rice, M. E., & Harris, G. T. (2003). *Planning services for a forensic mental health population.* Manuscript submitted for publication.

Rice, M. E., Harris, G. T., & Quinsey, V. L. (1990). A follow-up of rapists assessed in a maximum security psychiatric facility. *Journal of Interpersonal Violence, 5,* 435–448.

Rice, M. E., Quinsey, V. L., & Harris, G. T. (1991). Sexual recidivism among child molesters released from a maximum security psychiatric institution. *Journal of Consulting and Clinical Psychology, 59,* 381–386.

Sjöstedt, G., & Langström, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. *Psychology, Crime and Law, 8,* 25–40.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest: A Journal of the American Psychological Society, 1,* 1–26.

Webster, C. D. (1998). Comment on Thomas Mathiesen's selective incapacitation revisited. *Law and Human Behavior, 22,* 471–476.

Received September 16, 2002
Revision received April 1, 2003
Accepted April 8, 2003 ■